Stabilization and Rational Design of Serine Protease AprM under Highly Alkaline and High-Temperature Conditions

AKIHiko MASUI,1* NOBUAKI FUJIIWARA,1 and TADAYUKI IMANAKA2
Osaka Prefectural Industrial Technology Research Institute, 2-1-53, Enokojima, Nishi-ku, Osaka 550,1 and Department of Biotechnology, Faculty of Engineering, Osaka University, Yamadaoka, Suita, Osaka 565,2 Japan

Received 15 March 1994/Accepted 8 July 1994

Rational shift of the optimum pH toward alkalinity and enhancement of thermostability were investigated by using a thermostable extremely alkaline protease (optimum pH, 12 to 13) from the alkaliphilic and thermophilic Bacillus sp. strain B18'. The protease gene (aprM) was cloned, and the sequence analysis revealed an open reading frame of 361 amino acids that was composed of a putative signal sequence (24 amino acids), a prosequence (69 amino acids), and a mature enzyme (268 amino acids) (molecular weight, 27,664). The amino acid sequence of this protease was compared with those of other serine proteases. A direct correlation of higher optimum pH with an increase in the number of arginine residues was observed. An even more thermostable mutant enzyme was created by introducing a point mutation. When the position of the β-turn, Thr-203, was replaced by Pro, the residual activity of this mutant enzyme at 80°C for 30 min was higher than that of the wild-type enzyme (50% versus 10%). The specific activity of this mutant enzyme at 70°C was 105% of that of the wild-type enzyme under non-denaturation condition. These data suggest that the higher content of Arg residues favors the alkalinity of the serine protease and that introduction of a Pro residue into the β-turn structure stabilizes the enzyme.

Materials and Methods

Strains and plasmids. Strains and plasmids used in this study are listed in Table 1. Bacillus sp. strain B18' was used as a DNA donor. Escherichia coli JM109 was used as a host for the recombinant plasmid. Bacillus subtilis DB104, which is deficient in both neutral and alkaline proteases, and plasmid pISA412 were used as host and vector for the production of the cloned aprM gene. E. coli TG1 was used to subclone the gene for nucleotide sequencing. Plasmid pBR322 was used to construct the gene bank of Bacillus sp. strain B18' chromosomal DNA.

Media. All cultures were grown in L broth (7) with the exception of medium for Bacillus sp. strain B18'; it contained Na2CO3. Antibiotics (50 mg of ampicillin per ml or 25 mg of tetracycline per ml) were added to the media for cultivation of plasmid-harboring cells.

Preparation of plasmids and chromosomal DNA. Either the rapid alkaline extraction method or CsCl-ethidium bromide equilibrium density gradient centrifugation was used to prepare the plasmid DNA (11). Chromosomal DNA was prepared from a sarcosyl lysate of cells as described previously (15).

Southern blot and colony hybridization. In our previous work, we determined the sequence of the N-terminal 20 amino acid residues of the alkaline protease from Bacillus sp. strain B18' (5). The oligonucleotide 5'-CC(G/A/T/C)CTGGGG(G/A/T/C)AT(T/C)TG(A/T/C)TTT(T/C)AT(T/C)QA(T/C)AC-3' was used as a probe. This was based on the N-terminal amino acid sequence residues 4 to 12, Pro-Trp-Gly-Ile-Ser-Phe-Ile-Asn-Thr. It was labeled with [γ-32P]ATP and T4 polynucleotide

* Corresponding author. Mailing address: Osaka Prefectural Industrial Technology Research Institute, 2-1-53, Enokojima, Nishi-ku, Osaka 550, Japan. Fax: 81-6-443-3137.
kinase. The 0.4-kb HindIII-XhoI fragment was used as a random extension primer in nonradioactive DNA labeling. Southern blotting was performed as previously described (17).

**Gene library construction.** To construct the 1.4-kb HindIII library, total *Bacillus* sp. strain B18' DNA was digested with HindIII; 1.1- to 1.7-kb fragments were eluted from a 1.0% agarose gel and ligated to HindIII-digested pBR322. Similarly, a 3.0-kb Clal library was constructed by using Clal-digested pBR322. In all cases, the ligations were done with a 3:1 ratio of insert to vector. The ligation mixtures were incubated at 16°C overnight and used to transform *E. coli* JM109.

**Transformation.** Transformation of *E. coli* cells with plasmid DNA was performed by the method of Lederberg and Cohen (16). Transformation of *B. subtilis* competent cells with plasmid DNA was done as previously described (2).

**DNA sequencing.** DNA sequencing was done by the dideoxy-chain termination method with T7 DNA polymerase (19).

**Site-directed mutagenesis.** The synthetic oligonucleotide 5'-AGCACGTACCCAGGCAATC-3' was used for site-directed mutagenesis. Point mutations were introduced into genes by using an in vitro mutagenesis system (Amersham Co., Amersham, United Kingdom) as specified by the manufacturer.

**Protein assay.** The protein concentration was measured by the bicinchoninic acid protein assay method (27) with bovine serum albumin as the standard.

**Assay of alkaline protease.** Alkaline protease production was detected by comparing the size of the halo around colonies on LC broth (L broth supplemented with 1% casein) agar plates (30). Quantitative analysis of alkaline protease was done by the method of Amory et al. (1), and 1 U of enzyme activity was defined as the amount required to increase the A440 by 0.1 unit.

**Nucleotide sequence accession number.** The nucleotide sequence data reported in this paper have been entered into the GenBank, DDBJ, EMBL, and NCBI nucleotide sequence databases with the accession number D26542.

---

**RESULTS**

Cloning of an extremely alkaline protease-encoding gene, *aprM*, from *Bacillus* sp. strain B18'. The 26-mer probe, labeled with [γ-32P]ATP, was used for Southern blotting with digested chromosomal DNA of *Bacillus* sp. strain B18'. The digestion was done with several restriction enzymes. The probe hybridized to a 1.4-kb HindIII fragment (data not shown). A gene bank of 1.4-kb HindIII fragments based on size selection was constructed as described in Materials and Methods. When the labeled probe was used, 2 of 1,000 screened colonies were found to be positive in colony and Southern hybridization analyses. The two positive clones contained identical plasmids with three HindIII fragments (1.4 + 1.4 + 1.4 kb) (pABT40) (Fig. 1). Preliminary determination of the DNA sequence of the insert in plasmid pABT40 indicated that the HindIII fragment contained the N-terminal portion of the *aprM* gene but the C-terminal sequence was not included. To clone the entire *aprM* gene, a 0.4-kb HindIII-XhoI fragment was used as a nonradioactively labeled DNA probe and Southern blotting was performed by the procedure described above. Since the probe was hybridized to a 3-kbp Clal fragment (data not shown), a gene bank of 3-kbp Clal fragments was then constructed. In colony hybridization analysis, 3 of about 450 ampicillin-resistant colonies gave positive results. These three colonies carried identical plasmids. However, the inserted fragment was not a 3-kbp fragment as we predicted for pABT30 (Fig. 1). Therefore, we constructed a plasmid, pABT17, which lacked the Clal site. It excluded the 1.2-kbp region shown by the dotted line in pABT30 (Fig. 1). This plasmid (pABT17) contained an additional 300 bp at the C-terminal region from the HindIII site and included the whole *aprM* structural gene.

**Nucleotide sequence of the *aprM* gene.** The nucleotide sequence of the cloned protease gene and its flanking DNA regions was obtained (Fig. 2). DNA sequencing revealed an open reading frame (positions 602 to 1687 in Fig. 2). At 12 bases upstream from the start codon (ATG), there is a putative Shine-Dalgarno sequence, AGGAGG, which is complementary to the 3' terminus of the 16s rRNA from *B. subtilis* (20). This open reading frame encodes a polypeptide of 361 amino acids. Our previous determination of the amino acid sequences (5) showed that the N-terminal amino acid used to form the mature protein is Gln-94. In the proteins that are secreted, the signal peptide usually contains approximately 30 amino acids. This signal contains uncharged amino acids in the middle and positively charged amino acids close to its N terminus. In this open reading frame, the putative signal peptide contains 24 amino acids and possesses positively charged residues (Arg and Lys) followed by hydrophobic amino acids in the N terminus as reported previously (13). The prosequence contains an additional 69 amino acids and is a little shorter than that of other *Bacillus* serine proteases. The mature protease contains 268 amino acids, and the calculated molecular mass is 27,664 Da in good agreement with the previously determined value of 30,000 Da obtained by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The amino acid composition of the mature protease as determined by the nucleotide sequence is in good agreement with that previously determined by amino acid analysis (5).
Comparison of the amino acid sequence of AprM with those of other serine proteases. The amino acid sequences of AprM and other serine proteases (subtilisin BPN’ [33], subtilisin E [28], subtilisin Carlsberg [12], YaB elastase [15], and PB92 protease [31]) are compared in Fig. 3. The homology of the amino acid sequences of AprM and the other proteases was very high. The homology is about 70% for the entire open reading frame and 80% for the mature region. The Asp, His, and Ser residues which form the catalytic center of subtilisin-type proteases are conserved. It should be noted that there is a deletion of four consecutive amino acids in the substrate-binding pocket (4) as shown in both YaB elastase and alkaline PB92 protease.

Extremely alkaliphilic proteases have a high Arg content. The numbers of charged residues in the mature proteins of different serine proteases from Bacillus spp. are summarized in Table 2. By considering the Lys and Arg residues, which are positively charged, as well as the Glu and Asp residues, which are negatively charged, we found that AprM, PB92, and YaB contained fewer negatively charged residues than BPN’, E, and Carlsberg did. The total number of positively charged residues is not very different in these six proteases. When Arg and Lys are considered separately, it is seen that AprM contains 13 Arg and 2 Lys residues. Similarly, the numbers of Arg and Lys residues are 8 and 5 in PB92 protease, 9 and 4 in YaB elastase, 2 and 11 in subtilisin BPN’, 4 and 8 in E, and 4 and 9 in Carlsberg, respectively. Serine proteases with a higher optimal pH range tend to contain more Arg residues.

Enhancement of thermostability by amino acid substitution with Pro in the \( \beta \)-turn structure. The three-dimensional structures of subtilisin BPN’ and several homologous serine proteases have already been determined by X-ray crystallographic analysis (22, 32, 34). Overall folding patterns of these polypeptide chains closely resemble one another. In subtilisin BPN’, Gln-Ser-Thr-Leu (residues 206 to 209) form a \( \beta \)-sheet structure, Pro-Gly-Asn and Pro-Gly-Asp (residues 210 to 211) form a \( \beta \)-turn, and Lys-Tyr-Gly-Ala-Tyr (residues 212 to 213) form a reversed \( \beta \)-sheet. Pro-210 exists at the end of \( \beta \)-sheet (in other words, it is between the \( \beta \)-sheet and reverse \( \beta \)-sheet boxed position in Fig. 3).

From the viewpoint of steric structure and stability of entropy, Pro should be a more stable structure than Thr (3, 8, 18). Therefore, the stability of AprM should increase when Thr at the \( \beta \)-turn in AprM is substituted by Pro. Thr-203 in AprM (corresponding to Pro-210 of subtilisin BPN’) was substituted with Pro (T203P) by site-directed mutagenesis with a chemically synthesized oligonucleotide. The temperature effect on enzyme stability was determined by incubating the enzyme at
FIG. 3. Primary-structure homology of AprM with other Bacillus serine proteases. Ca, subtilisin Carlsberg. The residue numbers above the sequences are those of AprM, and the numbers below the sequences are those of subtilisin BPN'. Asterisks indicate the residues of the catalytic triad. The site-directed mutation position is boxed.

each temperature for 10 min (Fig. 4A). The enzymes were found to be stable from 50 to 70°C. The residual activities of this mutant enzyme are 75 and 6% at 80 and 90°C, respectively. On the other hand, the wild-type enzyme was completely inactivated at 90°C. To test its thermostability, the enzyme was kept in 0.1 M borate-NaOH buffer (pH 10.0) at 75 and 80°C for 30 min. The residual activities were then measured at 37°C. As shown in Fig. 4B, the activities of this mutant enzyme (T203P) are higher (80 and 50%) than those of the wild-type enzyme (50 and 10%) at 75 and 80°C, respectively. Specific activity of mutant enzyme was found to be 105% (at pH 10 and 70°C) of the wild-type level (data not shown).

TABLE 2. Charged amino acid content in the mature enzyme

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Optimum pH</th>
<th>No. of residues with:</th>
<th>Negative charge (Asp + Gln)</th>
<th>Positive charge (Lys + Arg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AprM</td>
<td>12-13</td>
<td>11</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>PB92</td>
<td>10.5-12</td>
<td>9</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>YaB</td>
<td>10.5-12</td>
<td>9</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>BPN'</td>
<td>8.5-10.5</td>
<td>15</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>E</td>
<td>8.5-10.5</td>
<td>14</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Ca</td>
<td>8.5-10.5</td>
<td>13</td>
<td>13</td>
<td>9</td>
</tr>
</tbody>
</table>

DISCUSSION

An alkaliphilic and thermophilic serine protease of the Bacillus sp. strain B18 has been cloned and characterized. We found an open reading frame that contains a signal sequence of 24 amino acids, a prosequence of 69 amino acids, and a mature protein of 268 amino acids. The signal and prosequence of this AprM are significantly shorter than those of serine proteases from other Bacillus spp. The prosequence is especially shorter by at least 10 amino acids than the other proteases such as subtilisin Carlsberg, E, BPN', YaB elastase, and PB92 protease, which contain 76, 77, 78, 83, and 83 amino acids, respectively.

The mature protease of AprM consists of 268 amino acids. AprM is highly homologous with the other Bacillus serine proteases. It is worth noting that four residues next to position 127 are deleted. This is similar to both PB92 protease and YaB elastase from alkaliphilic bacilli. This deletion is clearly different from those in other proteases of alkalophilic bacilli and neutrophilic bacilli. Gly-166 in subtilisin BPN' (corresponding to Asn-159 of AprM) is located at the bottom of the pocket for the PI site (cleavage site of substrate) and plays a critical role in determining their specificities. The pocket formed by Arg-155, Gln-156, and Asn-159 residues of AprM around the deletion region is bigger than those of other Bacillus serine proteases. The field of reaction (active site and substrate-binding site) may be sterically concealed. It is not likely that this deletion somehow raises the pH optimum.
We have developed a more probable explanation for the extremely alkaline pH optimum of this enzyme by comparing the amino acid composition of the mature enzyme with that of other proteases. There is an increase in the isoelectric points. The number of negatively charged Asp plus Glu residues is 11, 9, and 8 in AprM, PB92 protease, and YaB elastase, respectively, and 15, 14, and 13 in subtilisin BPN', E, and Carlsberg, respectively. It appears that the increase in the isoelectric points in alkaliophilic proteases occurs via a reduction in the number of Asp and Glu residues. There is also an increase in the pKₐ value of these enzymes. Generally speaking, the pKₐ value of an enzyme can influence the pH profile (24, 25). AprM has a higher pKₐ value because of an increase in the number of arginine residues (the pKₐ of the side chain of Arg is 12.5) and a decrease in the number of lysine residues (pKₐ, 9.7). There are 13 arginine and 2 lysine residues in AprM. By comparison, PB92 protease has 8 and 5, YaB elastase has 9 and 14, subtilisin BPN' has 2 and 11, E has 4 and 8, and Carlsberg has 4 and 9, respectively (Table 2). We speculate that many of these Arg residues in AprM reside on the surface of the molecule and enhance its stability under highly basic conditions. Therefore, it seems that the highly alkaliophilic proteases adapt to more highly alkaline conditions by altering their surface charge at higher pH.

Subtilisin-type proteases do not have an S-S bond and have several α/β protein scaffolds. Stabilization of the α-helix, β-sheet, and β-turn structural elements is very important for the stability of the entire global structure. Thr-203 of AprM does not exist at the hydrophobic core. In this case, stabilization of the β-turn structure after substitution of Thr-203 at the end of the β-sheet with proline should be expected. Introduction of a Pro into the β-turn structure through a site-directed mutation, resulted in an increase in the residual activity of the enzyme. Pro apparently stabilizes the β-turn structure because it causes a decrease in the overall entropy and thus forms a more rigid structure. In fact, the mutant enzyme showed lower (70%) and higher (105%) specific activities at 37 and 70°C, respectively, than the wild-type enzyme did.

Our results support the concept of designing new functional proteins through protein engineering: higher contents of Arg residues may shift the optimum pH to more alkaline levels, and introduction of a Pro residue at the end of a β-turn enhances thermostability as long as the conformation of the main chain and global structure are not drastically changed.

REFERENCES


